

## TRANSMITTAL OF APPEAL BRIEF (Large Entity)

Docket No.

ARC920000023US1

In Re Application Of: Kruelen, et al.



Application No.	Filing Date	Examiner	Customer No.	Group Art Unit	Confirmation No.
09/848,430	May 4, 2001	Laurie Anne Ries	48146	2176	7814

Invention: AN EFFICIENT STORAGE MECHANISM FOR REPRESENTING TERM OCCURRENCE IN UNSTRUCTURED TEXT DOCUMENTS

COMMISSIONER FOR PATENTS:

Transmitted herewith in triplicate is the Appeal Brief in this application, with respect to the Notice of Appeal filed on

September 29, 2005

The fee for filing this Appeal Brief is: \$500.00

- A check in the amount of the fee is enclosed.
- The Director has already been authorized to charge fees in this application to a Deposit Account.
- The Director is hereby authorized to charge any fees which may be required, or credit any overpayment to Deposit Account No. 09-0441
- Payment by credit card. Form PTO-2038 is attached.

**WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.**

Signature

Dated: 11/29/05

Frederick E. Cooperrider, Esq.  
Registration No. 36,769

McGinn Intellectual Property Law Group, PLLC  
8321 Old Courthouse Road, Suite 200  
Vienna, Virginia 22182

12/01/2005 HAL111 00000090 090441 09848430

01 FC:1402 500.00 DA

CC:

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to "Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450" [37 CFR 1.8(a)] on

(Date)

*Signature of Person Mailing Correspondence*

**HAND DELIVERED**

*Typed or Printed Name of Person Mailing Correspondence*

Appellants' Brief on Appeal  
S/N: 09/848,430



**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE  
BOARD OF PATENT APPEALS AND INTERFERENCES**

In re Application of

Kruelen, et al.

**Serial No.: 09/848,430**

Group Art Unit: 2176

Filed: May 4, 2001

Examiner: Ries, Laurie Anne

For: **AN EFFICIENT STORAGE MECHANISM FOR REPRESENTING TERM  
OCCURRENCE IN UNSTRUCTURED TEXT DOCUMENTS**

Commissioner of Patents  
Alexandria, VA 22313-1450

**APPELLANTS' BRIEF ON APPEAL**

Sir:

Appellants respectfully appeal the rejection of claims 1-25 in the Office Action dated June 29, 2005. A Notice of Appeal was timely filed on September 29, 2005.

**I. REAL PARTY IN INTEREST**

The real party in interest is International Business Machines Corporation, assignee of 100% interest of the above-referenced patent application.

**II. RELATED APPEALS AND INTERFERENCES**

There are no other appeals or interferences known to Appellants, Appellants' legal representative or Assignee which would directly affect or be directly affected by or have a bearing on the Board's decision in this appeal.

Docket ARC920000023US1

### **III. STATUS OF CLAIMS**

Claims 1-25, all of the claims presently pending in the application, stand rejected on prior art grounds.

Claims 1, 3, 5, 7, 9, 11, 13, 15, and 17-25 stand rejected under 35 USC §103(a) as unpatentable over US Patent 5,895,470 to Pirolli et al., further in view of US Patent Publication 2002/0165707 to Call. Claims 2, 6, 10, 14, and 16 stand rejected under 35 USC §103(a) as unpatentable over Pirolli/Call, further, further in view of US Patent 5,950,189 to Cohen et al. Claims 4, 8, and 12 stand rejected under 35 USC §103(a) as unpatentable over Pirolli/Call/Cohen, further in view of US Patent 6,401,088 to Jagadish et al.

The rejections for all claims are being appealed.

### **IV. STATUS OF AMENDMENTS**

An Amendment Under 37 CFR §1.116 was filed on August 29, 2005. In the Advisory Action dated September 15, 2005, the Examiner alleged that the combination of dependent claims, previously of record, into an independent claim, also previously of record, raises a new issue and declined entry of the Amendment.

Therefore, the claims in the attached Appendix consists of the claims as amended in the Amendment Under 37 CFR §1.111 submitted on April 21, 2005.

### **V. SUMMARY OF CLAIMED SUBJECT MATTER**

Appellants' invention, as disclosed and claimed in independent claim 1, is directed to a method of converting a document corpus containing an ordered plurality of documents (Figure 1 shows three documents in a document corpus exemplarily having three documents; final two lines on page 9) into a compact representation in memory of occurrence data (see Figure 7). A first vector (see ALLDATA vector in Figure 7) is Docket ARC920000023US1

developed for the entire document corpus. This first vector is a listing of integers corresponding to terms in the documents such that each of the documents in the document corpus is sequentially represented in the listing.

In dependent claim 18, the method of claim 1 further includes developing a dictionary (Figure 2) comprising the terms contained in the document corpus and associating, with each dictionary term, an integer to be uniquely corresponding to that dictionary term. These integers are those integers used in the first vector.

In dependent claim 19, the method of claim 18 further includes developing a second vector for the entire document corpus that indicates the location of each document's representation in the first vector (see STARTMARKER vector in Figure 7).

In dependent claim 2, the method of claim 18 further includes developing a third vector for the entire document corpus that provides a sequential listing of floating point multipliers, each floating point multiplier representing a document normalization factor (see MULT vector in Figure 7).

This method of breaking a corpus of a plurality of documents into one or more single vectors that represent all of the documents in the corpus is not taught or even suggested in the prior art of record, since, relative to the prior art currently of record, the present invention makes the contribution to the art of the novel technique that the entire corpus of documents be considered as a single entity containing data that can be organized in a sparse representation, rather than a plurality of documents, each respectively containing information potentially of interest.

As explained in the final sentence on page 13, the present invention is particularly advantageous when the corpus includes a million or more documents, the dictionary contains less than 32,000 terms, and each document contains less than a thousand words and has only one occurrence or a small number of occurrences of dictionary terms.

As explained in the penultimate sentence on page 14, in comparison with conventional multi-dimensional arrays or sparse matrix representations, the present invention reduces memory requirement by an order of magnitude, thereby alleviating the memory problem mentioned in the final paragraph on page 2 for large data sets.

## **VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL**

Appellant presents the following issues for review by the Board of Patent Appeals and Interferences:

### **ISSUE 1: THE REJECTION BASED ON US PATENT 5,896,570 TO PIROLI ET AL.**

Whether the rejection under 35 U.S.C. § 103(a) can be maintained for any of the claims, when none of the references currently of record address a document corpus having a precisely-defined order of documents, as required by the claim preamble, thereby inherently requiring that the principle of operation of the primary reference, as well as all secondary references, be modified in order to arrive at the present invention by a preliminary step to provide a precisely-defined order of documents, when none of the references currently of record makes any suggestion to provide this preliminary step of defining a precise ordering of documents, and when the concept of a single vector used throughout the entirety of a document corpus makes no sense without an agreed-upon ordering of documents;

### **ISSUE 2: THE MODIFICATION OF PRIMARY REFERENCE PIROLI BY SECONDARY REFERENCE CALL**

Whether the combination of Pirolli and Call is proper under 35 U.S.C. § 103(a), when the two references address two clearly distinguishable problems, thereby inherently being non-analogous art, when the resultant combination would still fail to satisfy the plain meaning of the language of even the independent claims, and when the rejection currently of record would contradict itself upon making this combination;

### **ISSUE 3: THE REJECTION BASED ON FURTHER MODIFYING PIROLI/CALL BY US PATENT 5,950,189 TO COHEN ET AL**

Whether the rejection under 35 U.S.C. § 103(a) can be maintained for claims 2, 6, 10, 14, and 16, when the secondary reference Cohen teaches a normalization indicating similarity between two documents, in contrast to a normalization within a document, Docket ARC920000023US1

thereby inherently precluding that the plain meaning of the claim language would be satisfied even if the modification were to be made; and

**ISSUE 4: THE REJECTION BASED ON YET FURTHER MODIFYING PIROLI/CALL/COHEN BY US PATENT 6,401,088 TO JAGADISH ET AL**

Whether the rejection under 35 U.S.C. § 103(a) can be maintained for claims 4, 8, and 12, when the rejection currently of record uses the wrong legal standard of review, when the secondary reference addresses a concept of a probability of an occurrence of a term rather than an actual occurrence of the term, and when the secondary reference relied upon makes no suggestion of the specific equation recited in the claim;

**ISSUE 5: ADDITIONAL DEFICIENCY**

Whether the rejection under 35 U.S.C. § 103(a) can be maintained for claims 3, 7, and 11, when none of the references currently of record teaches the claimed feature of sorting the data within each document itself so that identical unique integers are adjacent, and when the Examiner improperly relies upon the description for developing an alphabetical listing of terms rather than the description required by the plain meaning of the claim language.

**VII. ARGUMENTS**

**ISSUE #1: THE REJECTION BASED ON PIROLI**

Appellants believe that the Pirolli reference is clearly patentably distinguishable from the present invention, when viewed from the perspective of one having ordinary skill in the art, since the present invention addresses the problem of, as described in independent claim 1, “converting a document corpus containing an ordered plurality of documents into a compact representation in memory of occurrence data”, whereas the primary reference Pirolli does not have a precisely-defined order of documents and addresses the entirely different problem of “analyzing the topology, content and usage of collections of linked documents”, as clearly described at lines 18-19 of column 3, and elsewhere.

Docket ARC920000023US1

The Examiner clearly does not agree, alleging that the rejection currently of record correctly addresses the converting of a document corpus into a compact representation.

#### A. THE EXAMINER'S POSITION ON THE PIROLI-BASED REJECTIONS

Specifically, in paragraph 6 of the Office Action dated June 29, 2005, the Examiner alleges in several places that primary reference Pirolli addresses “*... converting, organizing, and representing in a computer memory a document corpus containing an ordered number of documents (See Pirolli, Column 7, lines 35-39).*”

It is noted that lines 35-39 of column 7 of Pirolli state: “*An SCA engine processes text Web pages by treating their contents as a sequence of tokens and gathering collection and document level object and token statistics (most notably token occurrence).*”

Thus, it appears that the Examiner somehow interprets the terminology:

- “... processes text Web pages ...”;
- “... treating their contents as a sequence of tokens ...”; and/or
- “... gathering collection and document level object and token statistics (most notably token occurrence)”

as somehow implying that the Web pages have a precisely-defined ordering at the Web site being analyzed and that this description is somehow related to the process defined in the preamble of the independent claims wherein is required the process of: “*A method of converting a document corpus containing an ordered plurality of documents into a compact representation in memory of occurrence data.*”

#### B. APPELLANT'S POSITION ON THE REJECTIONS BASED ON PIROLI

**First, the Examiner's position is flawed as a matter of law.**

Appellants submit that, as clearly described in MPEP § 2111.01, claim interpretation is constrained by the “plain meaning” of the claim language, as would be interpreted by one having ordinary skill in the art.

Appellants first submit that one of ordinary skill in the art would not agree that the Web site described in Pirolli has a precisely-defined order of the Web pages, absent some Docket ARC920000023US1

Appellants' Brief on Appeal  
S/N: 09/848,430

preliminary step that defines some type of arbitrary ordering to be used in accordance with the method of the present invention. Appellants also submit that, absent some defined ordering of document (e.g., Web pages), the concept of a single vector describing the contents of the Web site (e.g., the document corpus) makes no sense, since there would be no mechanism to determine which document is involved at various points in the single vector.

In contrast, the present invention starts out with a precisely-defined order of the documents in the document corpus, as clearly required in the preamble of the independent claims.

Second, Appellants submit that one of ordinary skill in the art would not interpret the description:

*“An SCA engine processes text Web pages by treating their contents as a sequence of tokens and gathering collection and document level object and token statistics (most notably token occurrence).”*

as addressing anything other than the object described by the plain meaning of this wording. That is, Appellants submit that, even if the Examiner attempts to broadly interpret the independent claim language so that the description at lines 35-39 of column 7 is considered by the Examiner as describing a “conversion” of the Web page text into a sequence of tokens as ultimately represented by statistics on token occurrence, such interpretation clearly fails to satisfy the plain meaning of the terminology “... into a compact representation in memory of occurrence data ....” in the claim preamble.

Therefore, Appellants submit that the rejection currently of record fails to meet the initial burden of a *prima facie* rejection because of these two fundamental deficiencies of the primary reference Pirolli.

**Secondly, the Examiner's position is flawed as a matter of fact.**

Appellants assume that the Examiner intends the Web site to be considered as a document corpus. However, Appellants submit that one of ordinary skill in the art would not consider that the description in the first sentence of the Abstract (“... from a collection of linked documents at a locality ...”) satisfies the requirement in the claim preamble that Docket ARC920000023US1

Appellants' Brief on Appeal  
S/N: 09/848,430

the Web pages are considered to be in any defined ordering. As explained at lines 54-57 of column 3, a “link” is merely the mechanism used for the user to move from one page to another, and would be completely arbitrary.

In order to satisfy the plain meaning of the language of the preamble, the Examiner has the initial burden of demonstrating that some type of pre-defined ordering of the Web pages, which Web pages are presumed to correspond to the documents in the document corpus, has been executed in Pirolli. Without addressing this aspect of the present invention of the pre-defined ordering of the documents in the document corpus, there would be little value in attempting to continue evaluation of the present invention, since a listing of term occurrences in the format of a single vector for the entire document corpus would have no meaning, since the correlation would be unknown between the information in the single vector and the document having that term information.

Appellants submit that the rejection currently of record makes no such attempt to consider whether there is an initial ordering of the Web site pages and, therefore, is inherently deficient.

Second, Appellants submit that, to one having ordinary skill in the art, the technique in Pirolli of developing statistics for the tokens parsed from the Web pages is an entirely different concept, as indicated by the description in Pirolli itself at lines 17-19 of column 1:

*“The present invention is related to the field of analysis and design of linked collections of documents, and in particular to categorization of documents in said collection.”* (emphasis by Appellants)

Similar wording occurs at lines 17-18 of column 3 of Pirolli:

*“A system for analyzing the topology, content and usage of collections of linked documents is disclosed.”* (emphasis by Appellants)

Thus, even Pirolli itself describes the technique therein as directed toward the “analysis” of the data accumulated by the search, not a preliminary conversion process.

In contrast, the present invention is directed to the conversion of the data in a document corpus “... into a compact representation in memory of occurrence data.”

Appellants note that the analysis methods described in Pirolli might very well be useful on the converted data preliminarily obtained by the method described in the present Docket ARC920000023US1

invention, in order to obtain similarities between documents and/or process similarities with a search query, but such additional processing to analyze the document corpus data is not being described in plain meaning of the claimed invention.

Therefore, Appellants secondly submit that the rejection currently of record has failed to meet the initial burden of a *prima facie* rejection for the independent claims by failing to heed the plain meaning of the description in the preamble that the method of the claimed invention addresses the conversion of the data in the document corpus "... into a compact representation in memory of occurrence data ....", since one of ordinary skill in the art would not consider that analyzing data to generate statistical data for occurrence of tokens within a document would be a conversion of data into a "compact representation of occurrence data".

ISSUE #2: THE MODIFICATION OF PRIMARY REFERENCE PIROLI BY  
SECONDARY REFERENCE CALL

Appellants believe that it is improper to modify primary reference Pirolli by secondary reference Call, given that the two references address two different problems, as clearly confirmed by their different classifications by the USPTO itself and because the primary reference Pirolli addresses an analysis of document contents and the secondary reference addresses the problem of a preliminary conversion of text data into a representation using integers.

The Examiner clearly does not agree.

**C. THE EXAMINER'S POSITION ON THE MODIFICATION OF PIROLI BY CALL**

In the rejection currently of record, the Examiner alleges that the modification of Pirolli by Call would be proper, as indicated exemplarily in the second full paragraph on page 3 of the Office Action dated June 29, 2005:

*"Pirolli does not disclose expressly developing a first uninterrupted listing of integers to correspond to an occurrence of terms in the document corpus. Call discloses developing an uninterrupted array of integers corresponding to an occurrence of terms*

Docket ARC920000023US1

Appellants' Brief on Appeal  
S/N: 09/848,430

*(See Call, Figure 1, element 135, and Page 3, paragraph 0029). Pirolli and Call are analogous art because they are from the same field of endeavor of processing electronic text data. At the time of the invention it would have been obvious to a person of ordinary skill in the art to include the array of integers corresponding to an occurrence of terms of Call with the method of Pirolli. The motivation for doing so would have been to permit more efficient execution of processing functions of the type typically performed by data processors (See Call, Page 1, paragraph 0010). Therefore, it would have been obvious to combine Call with Pirolli for the benefit of permitting more efficient execution of processing functions of the type typically performed by the data processors to obtain the invention as specified in claim 5, 9, and 13.”*

Thus, as best understood, it appears that the Examiner agrees that the primary reference Pirolli fails to teach or suggest developing an uninterrupted listing of integers corresponding to an occurrence of terms in the document corpus, understood as being the Web pages on a Web site being analyzed by the method in Pirolli.

It further appears that the Examiner considers that the motivation to modify the primary reference would be to obtain the benefits of processing data, as described in the secondary reference Call, in paragraph 0010 on page 1.

It further appears that the Examiner impliedly agrees that the primary reference Pirolli fails to provide any conversion of the data prior to the analysis procedure described therein, let alone a conversion into a single vector consisting of a listing of integers.

Finally, it further appears that the analysis of the rejection currently of record is inherently contradictory, since the analysis initially alleges that the primary reference performs a conversion but then concedes that the primary reference is deficient in performing the conversion described by the single claim limitation.

**D. APPELLANT'S POSITION ON THE MODIFICATION OF PRIMARY  
REFERENCE PIROLI BY CALL**

**First, the Examiner's position is flawed as a matter of law.**

Appellants first submit that the inherent inconsistency in the analysis currently of record, wherein the analysis initially alleges that a conversion occurs in the primary reference and then concedes that the conversion described by the single claim limitation is absent from the primary reference, precludes this analysis as meeting the initial burden of a *prima facie* rejection.

Second, Appellants submit that this inherent inconsistency also disqualifies this reference as a primary reference, since clearly, the principle of operation of the primary reference would have to change in order to modify it to satisfy the claim limitation. Such change in principle is precluded in the 1959 CCPA holding in *In re Ratti*, 270 F.2d 810, 123 USPQ 349, as clearly described in MPEP §2143.01: “*If the proposed modification or combination of prior art would change the principle of operation of the prior art invention being modified, then the teachings of the references are not sufficient to render the claims prima facie obvious.*”

Third, Appellants respectfully traverse that Pirolli and Call are properly combinable since, as described in MPEP § 2141.01(a), the criterion for combining references for evaluation under 35 USC §103(a) is that “*... the references must either be in the field of applicant's endeavor or, if not, then be reasonably pertinent to the particular problem with which the invention was concerned.*”

As pointed out later in that section under the subsection “**ANALOGY IN THE ELECTRICAL ARTS**”, even “*Reference to a SIMM for an industrial controller was not necessarily in the same field of endeavor as the claimed subject matter merely because it related to memories. Reference was not found to be in a different field of endeavor because it involved memory circuits in which modules of varying sizes may be added or replaced, whereas the claimed invention involved compact modular memories. Furthermore, since memory modules of the claims at issue were intended for personal computers and used dynamic random-access-memories whereas reference SIMM was*

*developed for use in large industrial machine controllers and only taught the use of static random-access-memories or read-only-memories, the finding that the reference was nonanalogous was supported by substantial evidence."*

Appellants submit that neither the primary reference Pirolli nor secondary reference Call addresses the problem being addressed by the present invention, as defined in the preamble of the independent claims, of providing a method to convert data in the documents of an ordered document corpus into a compact representation of occurrence data.

Pirolli clearly fails to convert text data on the Web pages into a representation of terms by either integers or floating point numbers, since it clearly addresses a data processing procedure, and Call clearly addresses a preliminary conversion of text data for a document and does not suggest extending that technique to an ordered document corpus wherein a single vector can be developed.

Moreover, the Examiner seems to recognize the inconsistency in the rejection currently of record by attempting to use as motivation to modify Pirolli the concept that the processing would be more efficient if the method in Call were to be incorporated. That is, it is noted that the rejection currently of record acknowledges that not even a single claim limitation of the independent claims is satisfied by the primary reference Pirolli. Appellants submit that this is clearly because Pirolli addresses the data processing of the analysis necessary to provide the statistics, not the preliminary conversion that the Examiner attempts to simply add to the data processing of Pirolli.

Therefore, clearly, the Examiner's motivation to modify Pirolli is inconsistent with the characterization that Pirolli teaches a conversion of the data in the document corpus into a compact representation of occurrence data.

Fourth, the Federal Circuit held in *In re Mills*, 916 F.2d 680, 16 USPQ2d 1430, 1990, recited at MPEP §2143.01: "*The mere fact that references can be combined or modified does not render the resultant combination obvious unless the prior art also suggests the desirability of the combination.*" (emphasis in MPEP itself).

Appellants submit that the rejection currently of record clearly violates this Federal Circuit guideline.  
Docket ARC920000023US1

In essence, the Examiner merely alleges that the motivation to modify the primary reference would be to obtain the benefit of having made the modification, clearly a circular reasoning based on hindsight. Again, Appellants point out that neither the primary reference Pirolli nor the secondary reference Call is addressing the problem of the present invention. Therefore, the Examiner is clearly using the present invention as a roadmap.

**Secondly, the Examiner's position is flawed as a matter of fact.**

Appellants submit that, even if Pirolli were to be combined with Call, the combination would not satisfy the plain meaning of the independent claims, since neither reference makes a preliminary step to define an order for the documents, thereby providing a basis to use a single vector for the entire document corpus. Because of this basic deficiency, the most that can be reasonably asserted is that Call would provide to Pirolli a preliminary conversion of data in the Web pages to be in integer format. Each page, however, remains as an isolated entity, so that the Web site (document corpus) remains as a collection of documents represented in an integer format, to now be processed as separate documents in accordance with the method described in Pirolli that includes developing a matrix, similar to the conventional methods discussed by Appellants in their background discussion.

Therefore, having made this combination of Pirolli and Call, there would still be no suggestion to represent the Web site contents as a single vector of information, either in its original format or in a converted integer format.

Finally, Appellants point out that, contrary to the Examiner's characterization that: "*Pirolli does not disclose expressly developing a first uninterrupted listing of integers to correspond to an occurrence of terms in the document corpus*", Appellants submit that Pirolli expressly teaches using a matrix method for the analysis, an entirely different concept from that of a single vector of terms in the document corpus.

Therefore, contrary to the Examiner's characterization, Appellants submit that Pirolli actually teaches against the element that the Examiner looks to secondary reference Call to accommodate.

ISSUE #3: THE REJECTION FOR CLAIMS 2, 6, 10, 14, and 16, BASED ON PIROLI/CALL, AS FURTHER MODIFIED BY COHEN

Appellants submit that, even if Cohen were to be combined with Pirolli/Call, the combination would still fail to satisfy the plain meaning of the claim language.

The Examiner clearly does not agree.

**E. THE EXAMINER'S POSITION ON THE MODIFICATION BY COHEN**

In paragraph 7 on page 7 of the Office Action, the Examiner alleges:

*"Pirolli and Call do not disclose expressly developing a third uninterrupted listing for the entire document corpus, the third uninterrupted listing containing a sequential listing of floating point multipliers, each floating point multiplier representing a document normalization factor for a corresponding document in the document corpus. Cohen discloses developing a normalized vector containing floating point multipliers (See Cohen, Column 11, lines 1-39). Pirolli, Call, and Cohen are analogous art because they are from the same field of endeavor of processing electronic text data. At the time of the invention it would have been obvious to a person of ordinary skill in the art to include the normalized vectors of Cohen with the method of Pirolli and Call. The motivation for doing so would have been to accurately identify the high matches of document terms and their values (See Cohen, Column 9, lines 28-30). Therefore, it would have been obvious to combine Cohen with Pirolli and Call for the benefit of accurately identifying the high matches of document terms and their values to obtain the invention as specified in claims 2, 6, 10, 14, and 16.*

Thus, it appears that the Examiner considers that normalization described in Cohen for use in a query matching would benefit Pirolli and/or Call.

**F. APPELLANT'S POSITION ON THE MODIFICATION BY COHEN**

**First, the Examiner's position is flawed as a matter of law.**

Appellants submit that even if Cohen were to be combined, the result would not satisfy the plain meaning of the claim language. Cohen addresses a comparison between Docket ARC920000023US1

documents. In contrast, the plain meaning of the claim language requires the floating point number represent a “document normalization factor”, an entirely different concept from a measurement of comparison between two documents.

The benefit of this aspect of the present invention is described in the fourth full paragraph on page 11 of the specification and has nothing whatsoever to do with comparison with another document. Rather, it relates to the conversion process of the term data into the sparse matrix representation format described by the present invention.

Furthermore, even if Cohen were to be somehow incorporated into Pirolli/Call, the plain meaning of the claim language of having a third uninterrupted listing has not been addressed in the rejection.

**Secondly, the Examiner's position is flawed as a matter of fact.**

Appellants submit that Pirolli already has a mechanism to measure comparison between two Web pages (e.g., line 66 of column 5 through line 2 of column 6 and lines 49-63 of column 7). Therefore, there is no need to further modify Pirolli for the rationale provided in the rejection currently of record.

**ISSUE #4: THE REJECTION BASED ON FURTHER MODIFYING  
PIROLI/CALL/COHEN BY JAGADISH**

Appellants submit that, even if Jagadish were to be combined with Pirolli/Call/Cohen, the combination would still fail to satisfy the plain meaning of the claim language.

The Examiner clearly does not agree.

**G. THE EXAMINER'S POSITION ON THIS REJECTION BASED ON JAGADISH**

In paragraph 8 on page 8 of the Office Action, the Examiner alleges:  
*“Pirolli, Call and Cohen do not disclose expressly that the normalization factor is the number of occurrences of a specific term in the document that represents the*  
Docket ARC920000023US1

*reciprocal of the square root of the sum of squares of all term occurrences in the document. Jagadish discloses calculating a normalization factor using an algorithm that can be refined to determine the number of term occurrences in a document (See Jagadish, Figure 6, and Column 8, lines 14-46). Pirolli, Call, Cohen and Jagadish are analogous art because they are from the same field of endeavor of processing electronic text data. At the time of the invention it would have been obvious to a person of ordinary skill in the art to include the normalization factor of Jagadish with the method of Pirolli, Call and Cohen. The motivation for doing so would have been to obtain a quick estimate of the number of times a particular substring, or term, occurs (See Jagadish, Column 1, lines 23-24)."*

Thus, it appears that the Examiner considers that the normalization factor of Cohen can somehow benefit if modified to incorporate the normalization factor described in Jagadish.

#### H. APPELLANT'S POSITION ON THIS REJECTION BASED ON JAGADISH

##### **First, the Examiner's position is flawed as a matter of law.**

Appellants submit first that the rejection currently of record uses an improper legal standard, since "... can be refined ...." is clearly a statement of hindsight.

Second, Appellants submit that the description above clearly defines two different normalization factors and that modification of the first normalization factor by the second clearly changes the principle of operation of the first, and is, therefore, prohibited in an obviousness analysis for the reason previously recited.

##### **Secondly, the Examiner's position is flawed as a matter of fact.**

Appellants submit that the mechanism in Pirolli to measure comparison between two Web pages inherently includes a determination of how many times a term appears. Therefore, there is no need to further modify Pirolli for the rationale provided in the rejection currently of record.

Moreover, Appellants submit that the Examiner's characterization that "... do not disclose expressly that the normalization factor is the number of occurrences of a specific

*term in the document....*" misrepresents the fact that the normalization factor in these references are directed to a comparison between references (e.g., an entirely different purpose).

**ISSUE 5: ADDITIONAL DEFICIENCY FOR THE REJECTION FOR CLAIMS 3, 7, AND 11**

Appellants submit that the description in paragraph 0051 on page 5 of Call fails to satisfy the plain meaning of the claim language.

The Examiner clearly does not agree.

**I. THE EXAMINER'S POSITION ON THIS REJECTION**

In the first full paragraph on page 6 of the Office Action, the Examiner alleges: "*Call also discloses rearranging, or sorting, in the first vector, the order of the unique integers within the data for each document so that the terms are in alphabetical order which would [cause] all identical unique integers to be adjacent (See Call, Page 5, paragraph 0051).*"

**J. APPELLANT'S POSITION ON THIS REJECTION BASED ON JAGADISH**

**The Examiner's position is flawed as a matter of fact.**

Appellants submit that the description in paragraph 0051 does not address a "first vector" and does not address relocating identical terms to be adjacent. Rather, it clearly addresses an alphabetization of terms, an entirely different concept. In contrast, this step is used in the present invention to allow the normalization calculation described in claim 4.

Moreover, simply placing terms in alphabetical order would not preserve the number of occurrences of that term and, therefore, this description in Call fails to satisfy the plain meaning of the claim language.

Appellants' Brief on Appeal  
S/N: 09/848,430

## CONCLUSION

In view of the foregoing, Appellants submit that claims 1-25, all the claims presently pending in the application, are clearly enabled and patentably distinct from the prior art of record and in condition for allowance. Thus, the Board is respectfully requested to remove all rejections of claims 1-25.

Please charge any deficiencies and/or credit any overpayments necessary to enter this paper to Assignee's Deposit Account number 09-0441.

Respectfully submitted,

Dated: 11/29/05



Frederick E. Cooperrider  
Reg. No. 36,769

McGinn Intellectual Property Law Group, PLLC.  
8231 Old Courthouse Road, Suite 200  
Vienna, VA 22182-3817  
(703) 761-4100  
Customer Number: 21254

## **VIII. CLAIMS APPENDIX**

Claims, as reflected upon entry of the Amendment Under 37 CFR §1.111 filed on April 21, 2005:

1. (Previously presented) A method of converting a document corpus containing an ordered plurality of documents into a compact representation in memory of occurrence data, said method comprising:

developing a first vector for said entire document corpus, said first vector being a listing of integers corresponding to terms in said documents such that each said document in said document corpus is sequentially represented in said listing.

2. (Previously presented) The method of claim 18, further comprising:

developing a third vector for said entire document corpus, said third vector comprising a sequential listing of floating point multipliers, each said floating point multiplier representing a document normalization factor.

3. (Previously presented) The method of claim 18, further comprising:

rearranging, in said first vector, an order of said unique integers within the data for each said document so that all identical unique integers are adjacent.

Appellants' Brief on Appeal  
S/N: 09/848,430

4. (Original) The method of claim 2, wherein said normalization factor is calculated as:

$NF = 1/ (\sum x_i^2)^{1/2}$  , where  $x_i$  is the number of occurrences of a specific term in said document, so that NF represents the reciprocal of the square root of the sum of squares of all term occurrences in said document.

5. (Previously presented) A method of converting, organizing, and representing in a computer memory a document corpus containing an ordered plurality of documents, said method comprising:

for said document corpus, taking in sequence each said ordered document and developing a first uninterrupted listing of integers to correspond to an occurrence of terms in the document corpus.

6. (Previously presented) The method of claim 19, further comprising:

developing a third uninterrupted listing for said entire document corpus, said third uninterrupted listing containing a sequential listing of floating point multipliers, each said floating point multiplier representing a document normalization factor for a corresponding document in said document corpus.

7. (Previously presented) The method of claim 19, further comprising:

for each said document in said document corpus, rearranging said unique integers so that any identical integers are adjacent.

8. (Original) The method of claim 6, wherein said normalization factor is calculated as:

$NF = 1 / (\sum x_i^2)^{1/2}$ , where  $x_i$  is the number of occurrences of a specific term in said document, so that NF represents the reciprocal of the square root of the sum of squares of all term occurrences in said document.

9. (Previously presented) An apparatus for organizing and representing in a computer memory a document corpus containing an ordered plurality of documents, said apparatus comprising:

an integer determining module receiving in sequence each said ordered document of said document corpus and developing a first uninterrupted listing of unique integers to correspond to an occurrence of terms in the document corpus.

10. (Original) The apparatus of claim 9, further comprising:

a normalizer developing a third uninterrupted listing for said entire document corpus, containing a sequential listing of floating point multipliers, each said floating point multiplier representing a document normalization factor for a corresponding document in said document corpus.

11. (Original) The apparatus of claim 9, further comprising:

a rearranger rearranging said unique integers so that any identical integers for each said document in said document corpus are adjacent.

Appellants' Brief on Appeal  
S/N: 09/848,430

12. (Original) The apparatus of claim 10, wherein said normalizer calculates said normalization factor as:

$NF = 1 / (S x_i^2)^{1/2}$ , where  $x_i$  is the number of occurrences of a specific term in said document, so that NF represents the reciprocal of the square root of the sum of squares of all term occurrences in said document.

13. (Previously presented) A signal-bearing medium tangibly embodying a program of machine-readable instructions executable by a digital processing apparatus to perform a method to organize and represent in a computer memory a document corpus containing an ordered plurality of documents, said method comprising:

developing a first uninterrupted listing of unique integers to correspond to the occurrence of terms in the document corpus.

14. (Previously presented) The signal-bearing medium of claim 25, wherein said method further comprises:

developing a third uninterrupted listing for said entire document corpus, containing a sequential listing of floating point multipliers, each said floating point multiplier representing a document normalization factor for a corresponding document in said document corpus.

15. (Previously presented) A data converter for organizing and representing in a computer memory a document corpus containing an ordered plurality of documents, for use by a data mining applications program requiring occurrence-of-terms data, said representation to be based on terms in a dictionary developed for said document corpus and wherein each said term in said dictionary has associated therewith a corresponding unique integer, said data converter comprising:

means for developing a first uninterrupted listing of said unique integers to correspond to the occurrence of dictionary terms in the document corpus and; and means for developing a second uninterrupted listing for said entire document corpus containing in sequence the location of each corresponding document in said first uninterrupted listing, wherein said first listing and said second listing are provided as input data for said data mining applications program.

16. (Original) The data converter of claim 15, further comprising:

means for developing a third uninterrupted listing for said entire document corpus, containing a sequential listing of floating point multipliers, each said floating point multiplier representing a document normalization factor for a corresponding document in said document corpus.

17. (Original) The data converter of claim 15, further comprising:

means for rearranging said unique integers so that any identical integers for each said document in said document corpus are adjacent.

Appellants' Brief on Appeal  
S/N: 09/848,430

18. (Previously presented) The method of claim 1, further comprising:
  - developing a dictionary comprising said terms contained in said document corpus;
  - and
  - associating, with each said dictionary term, an integer to be uniquely corresponding to said dictionary term, said uniquely corresponding integers being said integers comprising said first vector.
19. (Previously presented) The method of claim 1, further comprising:
  - developing a second vector for said entire document corpus, said second vector indicating the location of each said document's representation in said first vector.
20. (Previously presented) The method of claim 5, further comprising:
  - developing a dictionary comprising terms contained in said document corpus; and
  - associating, with each said dictionary term, an integer to be uniquely corresponding to said dictionary term, said uniquely corresponding integers used in said first uninterrupted listing.
21. (Previously presented) The method of claim 5, further comprising:
  - developing a second uninterrupted listing for said entire document corpus, said second uninterrupted listing containing, in sequence, the location of each corresponding document in said first uninterrupted listing.

22. (Previously presented) The apparatus of claim 9, further comprising:  
a dictionary developing module to develop a dictionary of terms contained in said document corpus, each said term being associated with a corresponding unique integer.

23. (Previously presented) The apparatus of claim 9, further comprising:  
a locator module developing a second uninterrupted listing for said entire document corpus, said second uninterrupted listing containing, in sequence, the location of each corresponding document in said first uninterrupted listing.

24. (Previously presented) The signal-bearing medium of claim 13, wherein said method further comprises:  
developing a dictionary comprising terms contained in said document corpus; and associating, with each said dictionary term, an integer to be uniquely corresponding to said dictionary term, said uniquely corresponding integers used in said first uninterrupted listing.

25. (Previously presented) The signal-bearing medium of claim 13, wherein said method further comprises:  
developing a second uninterrupted listing for said entire document corpus, said second uninterrupted listing containing, in sequence, the location of each corresponding document in said first uninterrupted listing.

Appellants' Brief on Appeal  
S/N: 09/848,430

**IX. EVIDENCE APPENDIX**

**(NONE)**

**X. RELATED PROCEEDINGS APPENDIX**

**(NONE)**